Problem Chosen	2021 MCM/ICM Summary Sheet	Team Control Number
D		2120499

Inheritance and Innovation of Music Genres

Music is not only an art form, but also witnesses the changes in human society. When creating new music, there are many factors that affect artists. Among them, the previously produced music is one of the most important factors. The purpose of this paper is to create models that capture ‘music influence’ and, through this perspective, quantify the evolution of music.

For the first question, in order to establish a graph theory model, after making full use of the data, we first defined two indicators that affect the weight and generated the weight adjacency matrix. Then, by using the Fruchterman-Reingold algorithm (FR), a directed network connecting influencers and followers was built, where weight is a parameter that capture 'music influence'. Subsequently, we selected a sub-network connecting specific artists. And we calculated the value of 'music influence' (weight) in this sub-network. Finally, based on the previous network, we built a new one with nodes arranged in order of time and genre to analyze the evolution of music clearly.

For the second question, we first used factor analysis to reduce the dimensions of 14 music characteristics into 3 main factors (intense, emotion and live). Then, cluster analysis was used to build a music similarity measurement model. In this process, the Gaussian Mixture Model (GMM) was used to model the probability distribution of each class and the Expectation Maximization (EM) Algorithm was used to maximize the GMM likelihood with respect to our artists data. Finally, we set up 5 clusters to represent different kinds of music. Utilizing this measurement model, we found that the similarity of the same music genre is higher than the similarity between different music genres.

For the third and fourth questions, the similarities and differences between types are mainly obtained by the influence coefficient between types. The influence coefficient is obtained by adding the influence of all artists in a type of music. The larger the influence coefficient is, the more similar it is. By observing the main music eigenvalues of a specific category, we found that the main music eigenvalues change with time. The relationship between types is mainly judged by the influence index and the similarity of music features.

For the fifth and the sixth questions, we analyzed the influence process in the evolution of Pop/Rock music from two aspects: the influence between genres and within the genre. With the help of the "out degree" in the graph theory model, we identified the representatives — The Beatles, Bob Dylan and The Rolling Stones who made a major leap and are the most influential artists in the evolution of pop rock music. Finally, we explained the style trend of Pop/Rock through three factors representing different characteristics of music.

For the last question, we had a further discussion on the influence of politics, society, and technology on music, and how to incorporate them into the model. Additionally, we analyzed the strengths and weaknesses of our model.

Keywords: Graph theory; Factor Analysis; EM Algorithm; Gaussian Mixture Model

Contents

1 Introduction

1.1 Problem Background

With the development of globalization and science and technology, cultural communication between different artists have become more frequent. Under this condition, when an artist creates music, he is often influenced by other artists and artistic genres. In addition, innate ingenuity, social conditions, musical instruments and tools, and personal experience will also affect music creation. When a new music genre appears or the existing music form undergoes major changes, it is often the social change or the communication of artists that make the music form a qualitative change. It can be said that it is the mutual influence of artists and major social changes that have promoted the development of music. Therefore, music is not only an art form, but also a witness to social changes.

In order to better understand how music evolves with social changes, we can develop a method to quantify music evolution. The mutual influence between different artists is measured by comparing the similarity of song characteristics. In this way, a network of artists and music genres can be established to intuitively understand how past music has influenced new music and music artists. After adding the ‘time’ factor, we can also understand the development and transformation of music genres. This line of thinking helps us understand the importance of the mutual influence of music to music revolution.

1.2 Restatement of the Problem

Considering the background information and restricted conditions identified in the problem statement, we need to solve the following problems:

- Create a directed network of musical influence, where influencers are connected to followers. Use parameters to capture *‘music influence’* in this network. Develop a subset of musical influence and describe it.
- Build a model to measure music similarity. Determine whether artists within genre are more similar than artists between genres.
- Compare influences between and within genres. Explain how genres change and how the influence processes of music evolute over time.
- Identify indicators to analyze that occurred over time in one genre and reveal the dynamic influencers.
- Explain how to express information about cultural influence of music and identify them within the network.

1.3 Our Work

We mainly build two models to solve the problem and use models to measure the influence and similarity of music. The specific work is as follows:

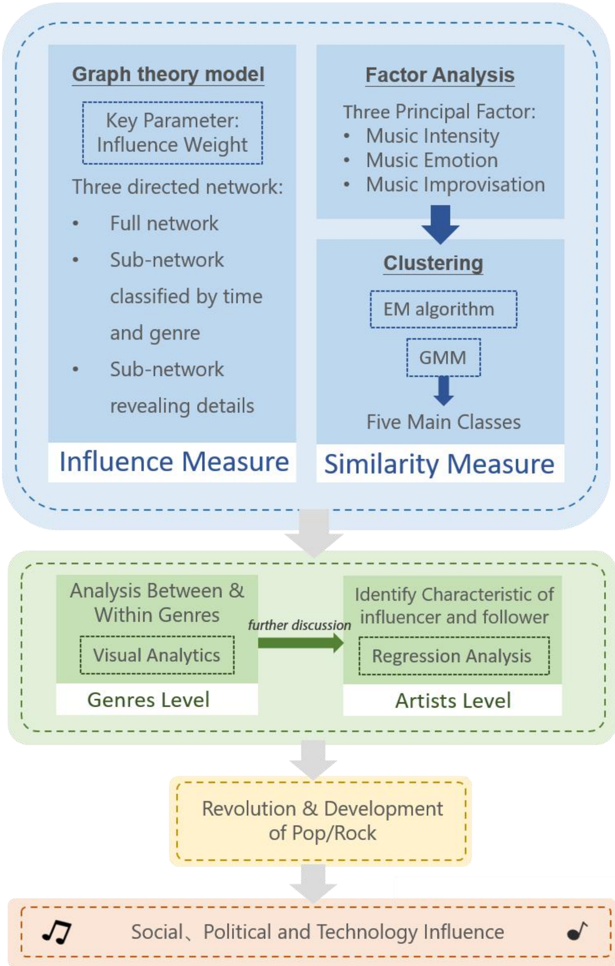

Figure 1: Flow chart of our work

2 Assumptions

- **Assumption1:** The time influence and genre influence have the same impact on weight
- **Assumption2:** All data are come from a few different gaussian probability distribution with different probability
- **Assumption3:** The music created by more than one artists belong to the genre of first artist

3 Notations

The key mathematical notations used in this paper are listed in Table 1.

Table 1: Notations used in this paper

Symbol	Description
M	music characteristic group
N	groups of data
D	distance between music features
K	group of clusters
ϕ	hidden variables

θ	parameters for GMM
$\rho(\varphi)$	function correlation probability distribution
α	the degree of correlation between music genres

4 The Directed Network of Musical Influence

4.1 Data Description

After descriptive statistical analysis of the *influence_data* data set, we get the following conclusions:

- There are 3774 influencers and 5046 followers. There are 5603 artists in total. Influencers and followers overlap.
- There are a total of 20 music genres. The following figure shows the number of followers of different music genres. It can be found that Pop/Rock, R&B, Country and Jazz account for a relatively large proportion and belong to the more mainstream music genres.

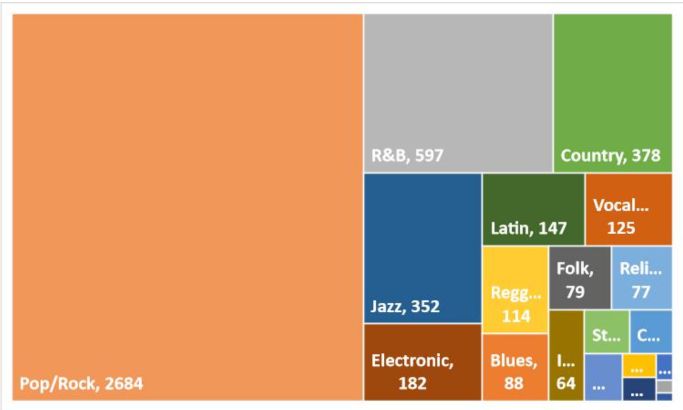

Figure 2: The number of followers of music genres

4.2 The Establishment of the Directed Network

4.2.1 The Weight Adjacency Matrix

Considering the relationship between followers and influencers, the directed network of the musical influence can be described as an adjacency matrix in Data Structure, which represents the connections between a node (follower) and an adjacency node (influencer). To generate an adjacency matrix, we determine indicators and define weight that characterize the connections. The indicators are as follows.

- **Time Gap** T **:** It means the time interval between the *influencer_active_start* and the *follower_active_start*. Generally, the smaller the time gap, the greater the impact. For the convenience of the adjacency matrix, we normalize them.
- **The degree of correlation between music genres** α : We define it by the number of artists when a specific genre is used as *follower_main_genre* and another specific genre is used as *influencer_main_genre*. It measures the degree of interconnection between genres. For example, among the artists belonging to Avant-garde, 11 artists are influenced by classical, so $\alpha = 11$ .The α of some genres are shown in the

following table2.

Table 2: The α of some genres

Followers' genre	Influencers' genres	α
Avant-Garde	Avant-Garde	7
	Classical	11
	Easy Listening	1
	International	2
	Jazz	2
	Pop/Rock	11
	R&B	1
	Stage & Screen	1
Children's	Children's	2
	Comedy/Spoken	1
	Jazz	1
	Stage & Screen	1
	Vocal	1

Then, we define α'.

$$\alpha'_{ij} = \frac{\alpha_{ij}}{\sum_{j=1}^{20} \alpha_{ij}} \tag{1}$$

i is followers' genre; j is influencers' genre

We use a heat map to show the value of α' as follows.

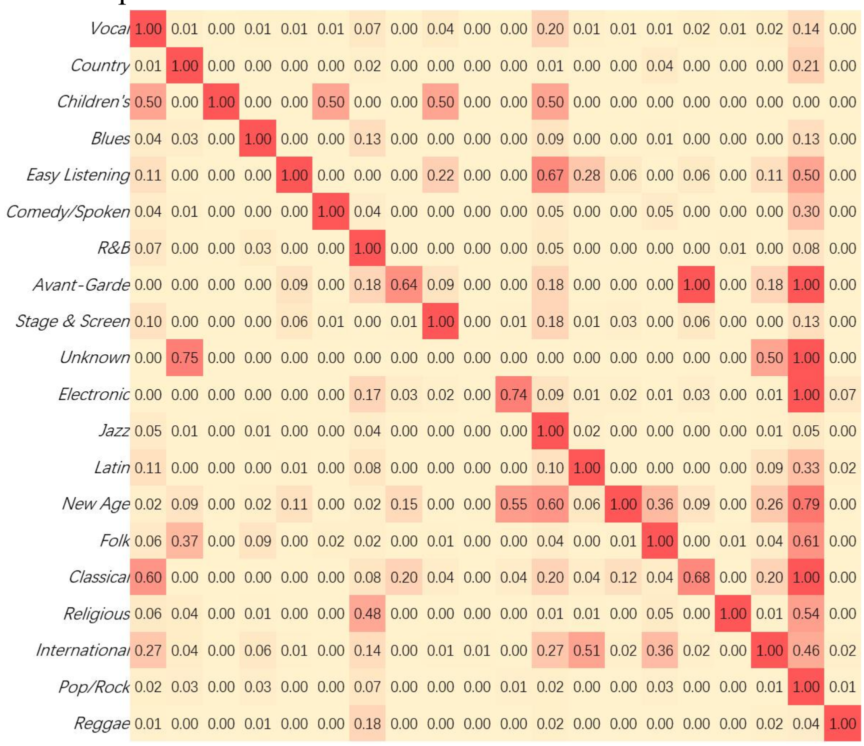

Figure 3: The degree of correlation between music genres

As shown in the figure, the abscissa is the followers' genres, and the ordinate is the influencers' genres. The values of each row represent the influence of influencers' genres on followers' genres. This influence can only be compared within one row. The order of genres on the abscissa is the same as the ordinate.

We define w_{ij} as the weight between influencer(i) and follower(j). It is also a critical parameter that captures 'music influence'. The calculation method is shown in Equation 1. If there is no influence between the two artists, the weight is 0.

$$w_{ij} = \frac{T_{ij} + \alpha'_{mn}}{2} \tag{2}$$

T_{ij} represents the positive and normalized time gap between i and j. α'_{mn} represents the value of correlation degree between music genre m and n.

The weight adjacency matrix between influencers and followers can be expressed as:

$$\begin{pmatrix} w_{11} & w_{12} & \cdots & w_{1,5046} \\ w_{21} & w_{22} & \cdots & w_{2,5046} \\ \vdots & \vdots & \cdots & \vdots \\ w_{3774,1} & w_{3774,2} & \cdots & w_{3774,5046} \end{pmatrix} \tag{3}$$

4.2.2 The Whole Directed Network

Based on the weight adjacency matrix, we use python and Fruchterman-Reingold algorithm(FR) to obtain a directed graph network between influencers and followers. FR is an improved energy model. In this model, the amount of energy is related to the weight. The greater the weight, the greater the energy. As shown in the figure4 , nodes of the same color represent the same genre.

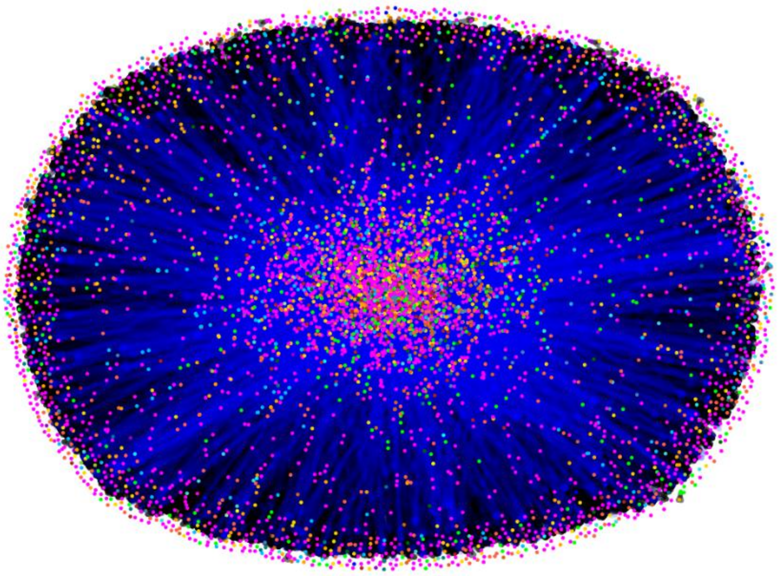

Figure 4: Directed network between influencers and followers

In order to illustrate more clearly, we arranged all artists in a coordinate system according to the genres and the decade that the artists began their music career. In this way, the artists' nodes of the same genre and decade overlap and are located on the same coordinate.

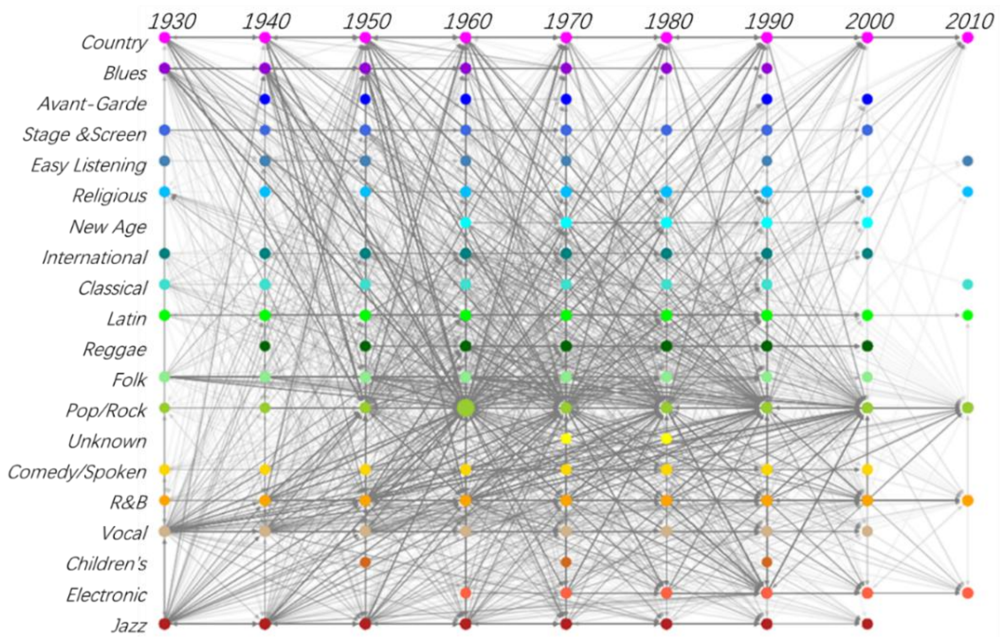

Figure 5: Directed network arranged in a coordinate system

The nodes represent all artists. The size of the node depends on the out-degree. Out-degree is the number of followers that an influencer influences, that is, the ability to influence. The larger the node, the greater the influence. Moreover, the width of the edges connecting the nodes represents the weight. Arrows point from influencers to followers. In graph theory, the ratio between the number of nodes and the number of edges is density. The density of this network is 0.00138.

4.3 Analysis of a Specific Subnetwork

4.3.1 Selection of the Subnetwork

Taking the number of followers influenced by each artist as the ability of influence, we randomly selected a sub-network from the fully connected network of the top 800 most influential people. This subnetwork is shown in the figure6.

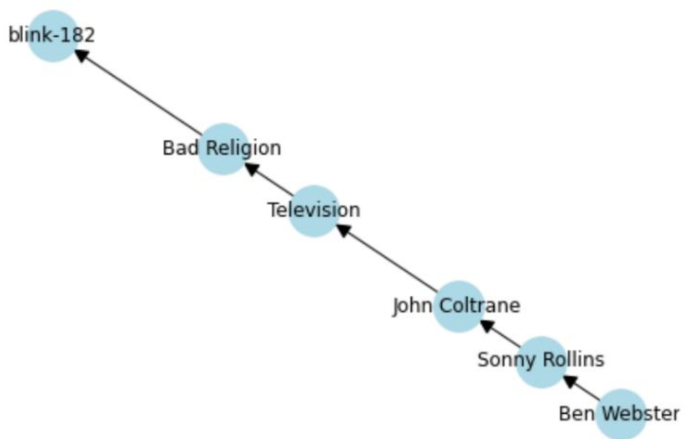

Figure 6: A specific directed subnetwork

4.3.2 Description of the Subnetwork

Based on Equation1, we calculate the influence of influencer on follower, which is equal to the weight. The results are shown in the table3.

Table 3: The influence in the subnetwork

Influencer 【gerne, year】	followers	influence
Bad Religion 【'Jazz', 1930】	blink-182 【'Pop/Rock', 1990】	0.75
Television 【'Pop/Rock', 1970】	Bad Religion 【'Jazz', 1930】	0.75
John Coltrane 【'Jazz', 1940】	Television 【'Pop/Rock', 1970】	0.33
Sonny Rollins 【'Jazz', 1940】	John Coltrane 【'Jazz', 1940】	0.71
Ben Webster 【'Jazz', 1930】	Sonny Rollins 【'Jazz', 1940】	0.75

5 Music Similarity Measurement Model

5.1 The Establishment of the Measurement Model

5.1.1 Factor Analysis

There are 12 feature variables (excluding mode and explicit variables). The reason for removing the two variables is that Boolean variables can't be extracted directly and affect the accuracy of factor analysis.

After the standardization, firstly, we calculate the factor load matrix. Factor load matrix is not sole, we estimate it by principal component analysis.

Secondly, we calculate the rotating factor load matrix. After rotation, the load of variables in each factor becomes either larger or smaller, which makes the common factor more explanatory. It can be shown in the figure7.

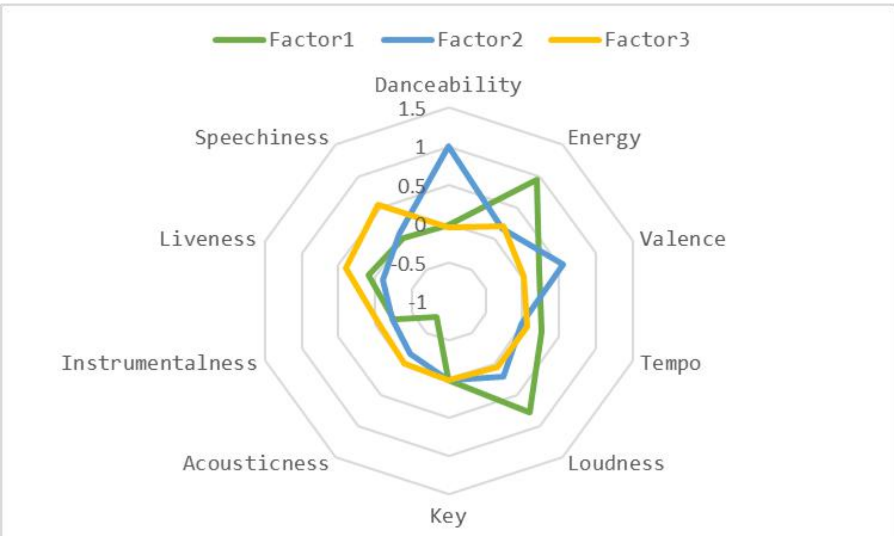

Figure 7: Radar chart of rotation factor loading matrix

Thirdly, we calculate the contribution rate of factor variance. The Aggregate contribution rate of $F_j(j = 1,2,3)$ to X_i equals to the sum of squares of the elements in the rotated factor load matrix, which can be illustrated as:

$$V_j = \sum_{i=1}^{3} l_{ij}^2 \tag{1}$$

It can be used to measure the relative importance of F_j, which means that contribution rate of factor variance can be portrayed as the weight of factors.

According to the musical characteristics of the three main factors, we named the three factors as music intensity, music emotion and music improvisation.

5.1.2 Cluster Analysis

Based on the result of factor analysis, in order to measure the similarity between the music, we use the cluster analysis model to explore the similarity of the whole music, and use the cosine similarity model to measure the similarity between the two music.

The cluster analysis model obtains the point distribution of different music through the maximum expectation algorithm, and judges the similarity between types. Its main principle is to judge the similarity by minimizing the distance of feature variables to achieve the proximity distribution.

Three indicators represent music characteristics, that is, M represents music, M_1, M_3, M_3 represents the three variables

$M = \{m_1, m_2, m_3\}$ data set contains N groups of data.

$$D(M,G) = \sum_{i=1}^{3} (m_i^T - g_i^T)^2 \tag{1}$$

Among them, D represents the distance between music features, and the closer the distance is, the closer it is; G represents a group of feature points of a group, and the smaller the distance is, the closer it is to the group feature. We're finally going to achieve d minimum, which is

$$E = \sum_{j=1}^{N} \sum_{i=1}^{3} D\,(M,G) \tag{2}$$

To solve the objective equation, we use the expected maximum solution. Given music data are independent of each other, $M_1 = [m_1, m_2, m_3]$, a probability model $f(M, \varphi, \theta)$ with hidden variables φ and parameters θ. According to MLE theory, the optimal single point estimation of parameter θ is given as $\theta = argmaxp(M/\theta)$ when the likelihood of the model is maximized.

Consider the case of implicit variables and assume discrete implicit variables. Then it has the following expansion.

$$p(M \mid \theta) = \sum_{u=1}^{k} p\,(M, \varphi_i \mid \theta), \varphi = \{\varphi_1, \varphi_2 ... \varphi_k\} \tag{3}$$

Take the natural logarithm of the above formula

$$\log p\,(M \mid \theta) = \log \prod_{i=1}^{n} p\,(M \mid \theta) = \sum_{i=1}^{n} \log p(M \mid \theta) = \sum_{i=1}^{n} \log \left[\sum_{i=1}^{K} p\,(M_i, \varphi_u \mid \theta) \right] \tag{4}$$

Hidden variables represent missing data or random variables that cannot be observed directly. Assuming that the hidden variable is a discrete variable, the natural logarithm of the above formula is taken. The implicit function correlation probability distribution $\rho(\varphi)$ is logarithmic likelihood and has the following unequal relation[1]

$$\log p\,(M \mid \theta) = \sum_{i=1}^{n} \log \left[\sum_{u=1}^{K} \frac{\rho(\varphi)}{\rho(\varphi)} p(M_i, \varphi_i \mid \theta) \right] \geq \sum_{i=1}^{n} \sum_{u=1}^{k} \left[\rho(\varphi_u) \log \frac{p(M, \varphi_i \mid \theta)}{\rho(\varphi_u)} \right] = L(\theta, \varphi) \tag{5}$$

EM algorithm is used to solve the objective function $\hat{\theta} = argmaxp(M/\theta)$. The EM algorithm is an iterative algorithm that finds a locally optimal solution $\hat{\theta}$ to the GMM likelihood maximization problem.

The EM algorithm consists of two steps.

- **E step:** It involves finding the posterior probability that point $\mathbf{x}^{(i)}$ was generated by cluster j, for every $i=1,...,n$ and $j=1,...,K$. This step assumes the knowledge of the parameter set θ. We find the posterior using the following equation:

$$p(point\,\mathbf{x}^{(i)}\,was\ generated\ by\ cluster\ j\mid\mathbf{x}^{(i)},\theta)\triangleq p(j\mid i)=\frac{p_j\mathcal{N}(\mathbf{x}^{(i)};\mu^{(j)},\sigma_j^2 I)}{p(\mathbf{x}^{(i)}\mid\theta)}. \tag{6}$$

- **M step:** It maximizes a proxy function $n\hat{\ell}(\mathbf{x}^{(1)},...,\mathbf{x}^{(n)}\mid\theta)$ of the log-likelihood over θ, where

$$\hat{\ell}(\mathbf{x}^{(1)},...,\mathbf{x}^{(n)}\mid\theta)\triangleq\sum_{i=1}^{n}\sum_{j=1}^{K}p\,(j\mid i)\log\left(\frac{p(\mathbf{x}^{(i)}\text{ and }\mathbf{x}^{(i)}\text{ generated by cluster }j\mid\theta)}{p(j\mid i)}\right). \tag{7}$$

This is done instead of maximizing over θ the actual log-likelihood

$$\ell(\mathbf{x}^{(1)},...,\mathbf{x}^{(n)}\mid\theta)=\sum_{i=1}^{n}\log\left[\sum_{j=1}^{K}p\,(\mathbf{x}^{(i)}\ generated\ by\ cluster\ j\mid\theta)\right]. \tag{8}$$

Maximizing the proxy function over the parameter set θ, one can verify by taking derivatives and setting them equal to zero that

$$\widehat{\mu^{(j)}}=\frac{\sum_{i=1}^{n}p\,(j\mid i)\mathbf{x}^{(i)}}{\sum_{i=1}^{n}p\,(j\mid i)}. \tag{9}$$

$$\widehat{p_j}=\frac{1}{n}\sum_{i=1}^{n}p\,(j\mid i). \tag{10}$$

The E and M steps are repeated iteratively until there is no noticeable change in the actual likelihood computed after M step using the newly estimated parameters or if the parameters do not vary by much.

Finally, the distribution map is constructed according to GMM. The characteristic variable of music is represented by Gaussian mixture model as follows:

$$P(x)=\sum_{k=1}^{K}\pi_k\,N(x\mid\mu k,\Sigma k). \tag{11}$$

Where $N(x\mid\mu k,\Sigma k)$ is the k-th component in the mixed model, there should be 5 clusters(we set up 5 groups based on our assumption) to represent different kinds of music. π_k is the mixing coefficient as the weight, and satisfies the

$$\sum_{k=1}^{K}\pi_k=1. \tag{12}$$

The parameters of Gaussian mixture model are calculated by EM algorithm, which makes the final algorithm converge and get the final parameters to form a clustering graph.

In addition, for the similarity between the two specific music, we established cosine correlation model as a supplement.

The different eigenvalues of each music constitute the feature vector of the music

$$M_1 = [m_1, m_2, m_3, m_4 ...]. \tag{13}$$

$$p(M_1, M_2) = \frac{\sum M_{1i} M_{2i} - n\overline{M_1 M_2}}{\sqrt{n\sum M_1^2 - \left(\sum M_1\right)^2}\sqrt{n\sum M_2^2 - \left(\sum M_2\right)^2}}. \tag{14}$$

This value can show the similarity between different music, the closer to 1, the more similar.

Through cluster analysis, we can observe the similarity degree of 20 music types when they are clustered into 5 categories. As shown in the figure8 below, we can roughly observe that different groups exist in different positions as clusters, so we can speculate that music similarity exists between categories.

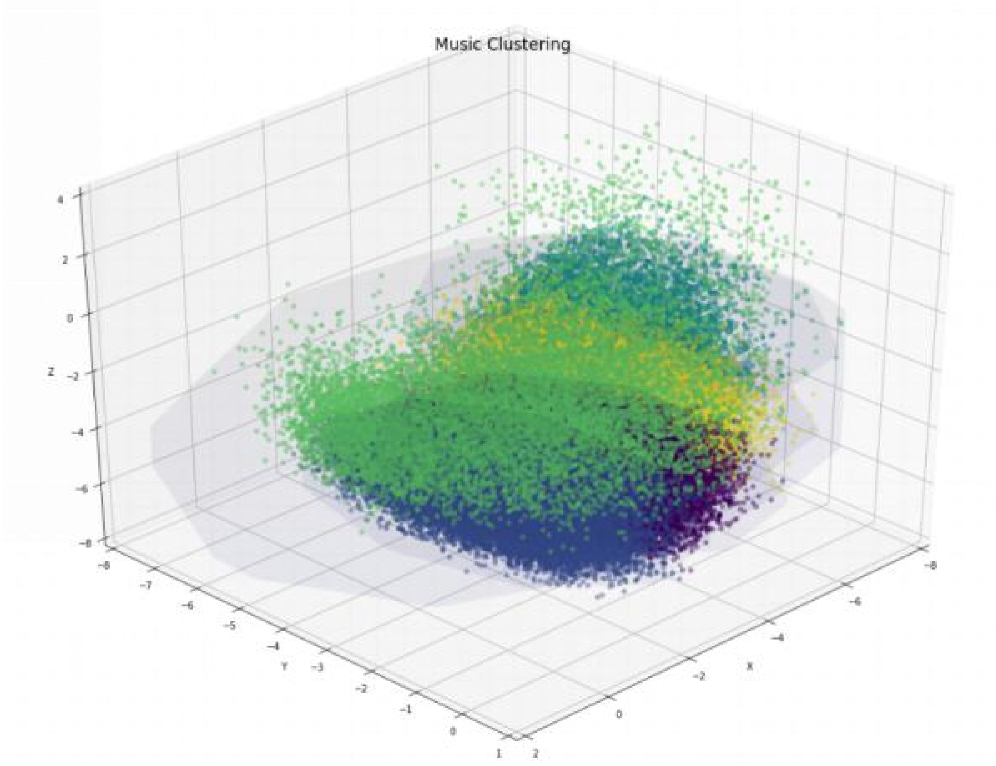

Figure 8: The result of the cluster analysis

5.2 Comparison of similarities within and between music genres

Through the presentation of GMM, we find that artists of the same music type have stronger similarities, and there are similarities and differences between different music types.

5.2.1 Similarity Measurement between Music Genres

According to the clustering results, the horizontal axis represents the clustering level, the vertical axis represents different kinds of music, and the size of the circle represents the percentage of artists in this field. The larger the percentage, the more concentrated the artists in this field are in this clustering level. The percentages of different types of musicians in a certain cluster level can reflect similarities and differences. For example, we can find that Folk music and Vocal music have great similarities. Children and Comedy music also have great similarities. Classical and New Age types have great similarities. Avant-Garde and Unknown are similar. Moreover, there is a great similarity between blues and reggae music to some extent.

Through horizontal and vertical comparison, we can also find that the similarity of artists of the same category is higher than that of different categories. For example, Avant-Garde artists mainly gather in type3, and New Age artists mainly gather in type2, which means that artists of the same category have a strong clustering effect. At the same time, the figure can also reveal that music types are related and have influence on each other.

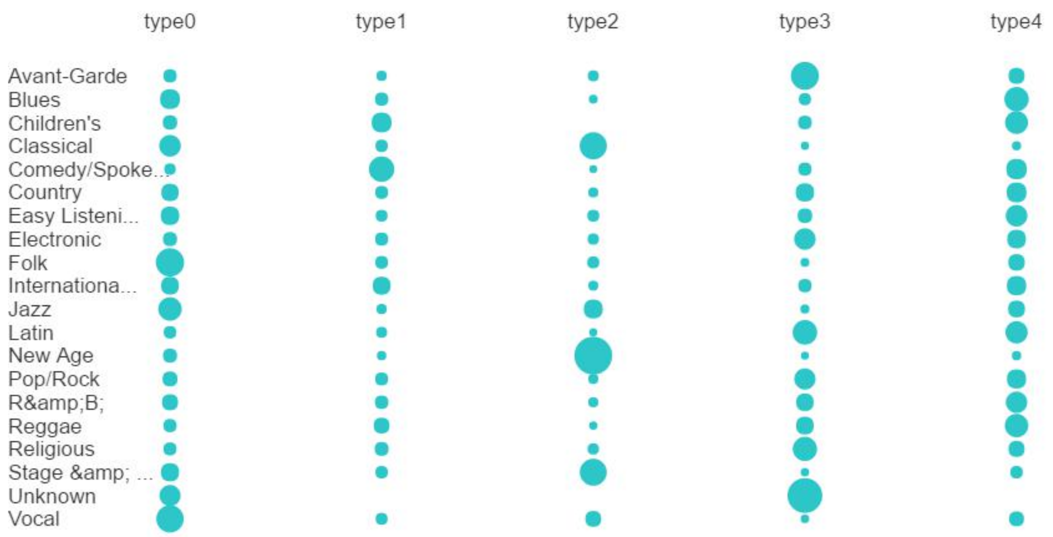

Figure 9: Similarity between genres

5.2.2 Similarity Measurement within Music Genres

In the same type of music, the percentage of artists at different aggregation levels reflects the difference of the same kind. Among them, pop is the most prominent one. According to our analysis of pop, nearly 50% of the artists cluster to 3, which is the area represented by the orange column, followed by 2 and 4 levels, which means that the commonness of pop exists in 3 levels, and the difference also exists in this music type. So, on the whole, there are some differences among pop types. The cluster distribution of Avant-Garde、Religious、Electronic music types is relatively consistent, which indicates that there are great differences among these music types. On the contrary, the differences between jazz and vocal music types are small.

Combined with the above two figures, artists of the same music type gather more. We conclude that the music similarity of the same music category is higher than that of different music categories.

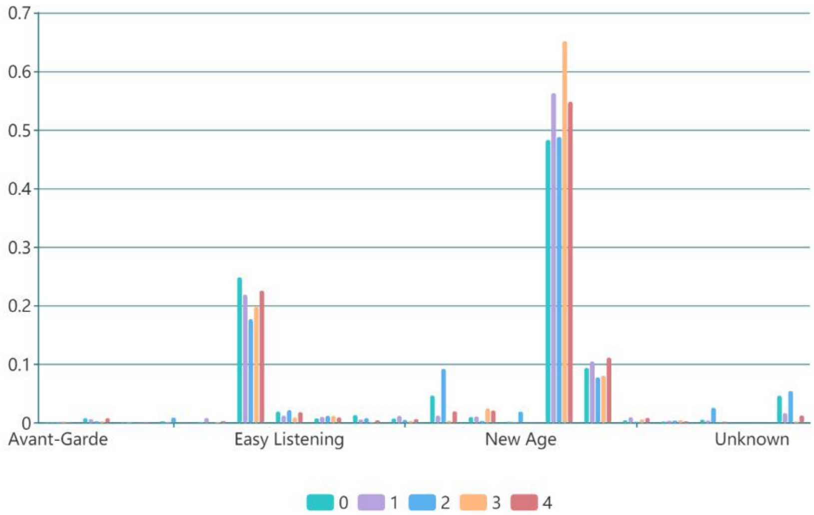

Figure 10: Similarity within genre

6 Quantify Music Influence and Characteristics

6.1 Influences Between and Within Genres

we consider the time, and sum up the influence of artists in the same year of the same music total influence in one genre in one year $= \sum_{indi=1}^{N} x_{influence}$. $x_{influence}$ represents the influence of artists.

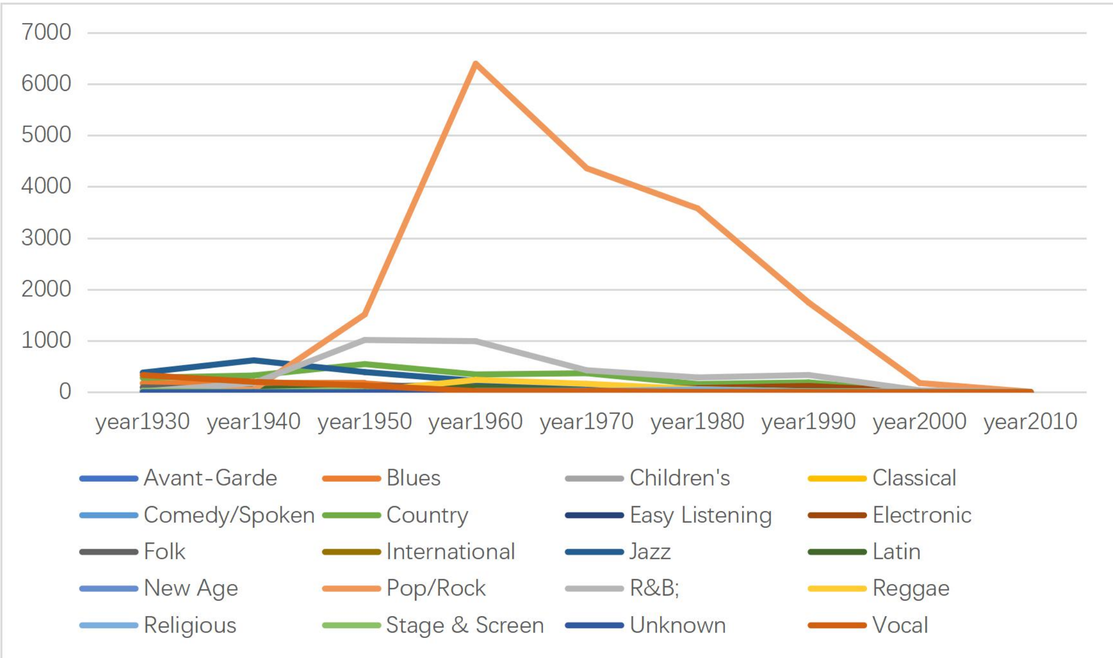

Figure 11: influence change over time

The figure11 above shows the changes in the influence of music genres, from which we can draw conclusions. Pop, as the most influential music type, gradually showed its influence in 1940, reached its peak in 1960, and then gradually declined. According to the time stage, avant garde was the most influential music type in 1940, followed by pop and R & B. The influence of Regge music increased significantly in 1960.

6.2 Gerne Characteristics and Changes over Time

Next, we make a more detailed analysis, a type of music internal eigenvalue changes with time using bubble chart. The size of the circle represents the aggregation of eigenvalues, which directly reflects the representativeness of music eigenvalues. Taking Avant-Garde music as an example, seven features are selected for analysis. We can conclude that danceability is a prominent feature of this type of music, and it is still the main characteristic of this genre of music since its birth, The music enthusiasm represented by valence gradually changes from positive to negative, and the speechiness also gradually decreases, which means that the existence of spoken language is weakened. The duration has similarity between different years and remains at a low value.

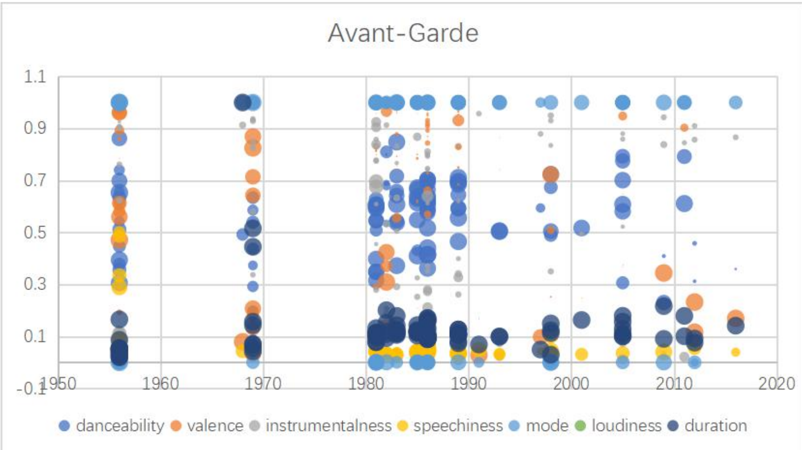

Figure 12: Characteristics change in gerne Avant-grade

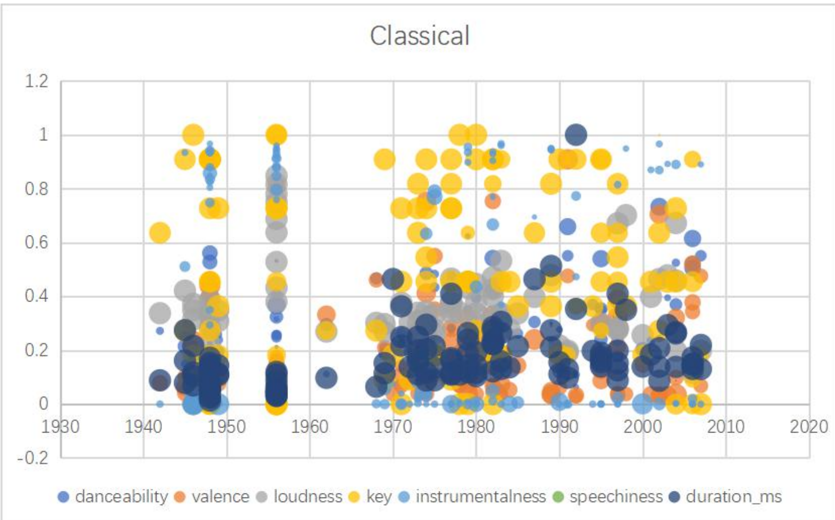

Figure 13: Characteristics change in classical

In order to compare the types, we make a typical feature map for analysis. The results show that the value of classical music is always at a low level, which indicates that its enthusiasm is not high, and the duration value is not very high, which is gathered around 0.2. We can infer that the characteristic indexes such as duration and loudness, which belong to intensity, can be used as the characteristic values of classical.

Taking pop music as an example, as one of the most influential types, it has a significant impact on music of different factions, with an impact of 1 on Avant-Garde, unknown, electronic and classic, and 0.79 on new age, which also has an impact on other types of music.

6.3 The Prove that Influencers Influence Followers

If there are different musical characteristics, it means that the influencer does influence. If the difference of musical characteristics is not obvious, it means that the influencer has little influence.

For the same influencer influencing different artists, the music characteristics of the affected artists are extracted and averaged to get the average music characteristics of the followers influenced by the artist. By comparing the average music characteristics of the same kind of music affected by different people, if the difference is large, it shows that the influencers do affect the music of the followers. If the sum of the absolute values of the differences between the average music feature values is greater than 10, it means that there is a great influence.

For example, 355 and 441 artists have influenced some musicians in pop category, and average the music characteristics of the musicians influenced by the same musician,$C_1=[C,C_2,C_3,...C_n]$ the array represents n values of music features.

The result shows that the difference is 13.9, and the default value is 10, which proves that the affected artists have influence. Due to the large influence of the deviation caused by the duration index, in order to simplify the processing, we remove the duration index.

6.4 Contagious Musical Characteristics

If different music features are subtracted between the influencer and the affected as x, and the influence of the two as y, a regression equation is constructed between x and y.

$$y=\beta_1x_1+\beta_2x_2+...\beta_nx_n+\epsilon\,. \tag{1}$$

ϵ represents error.

The coefficient before x can represent the musical feature. The larger the coefficient is, the higher the infectivity of the feature is. Different artists have different characteristic coefficients, which represent the characteristics of an artist. By comparing these coefficients, we can get the similar effect of a certain factor on the influence of artists.

7 Quantify the Evolution of Music

7.1 Revolutionaries Characteristic in Pop/Rock's Evolution

We use the number of followers to represent the influence of the influencer. Through grouping statistics, we find out the most influential influencers of the era from 1950, 1960, 1970, 1980, 1990, and 2000, and obtain all the influencers issued by them. Mean data of the characteristics of music. Correlation analysis of the musical characteristics of the 6 most influential artists in their respective eras, and the following correlation coefficient matrix which reveal the different among music characteristics.

Figure 14: Interaction of genres

7.2 Explore Evolution of Pop/Rock Music

The evolution of music genres will be influenced by both previous artists of the same genre and artists of others genres. In the process of evolution, representatives will continue to make huge leaps, leading a genre to be more prosperous. Since the Pop/Rock is one of the most popular music genre, we will discuss the representatives make revolutions or develop Pop/Rock music and the influence of other music genres on its evolution.

7.2.1 Influence within Pop/Rock Genre

From the figure3 denoting the degree of correlation between music genres we got from question 1, We can clearly observe that the development of Pop/Rock music is mainly influenced by artists from the same genre. We searched our network for the number of influencers in different eras and got the following figure：

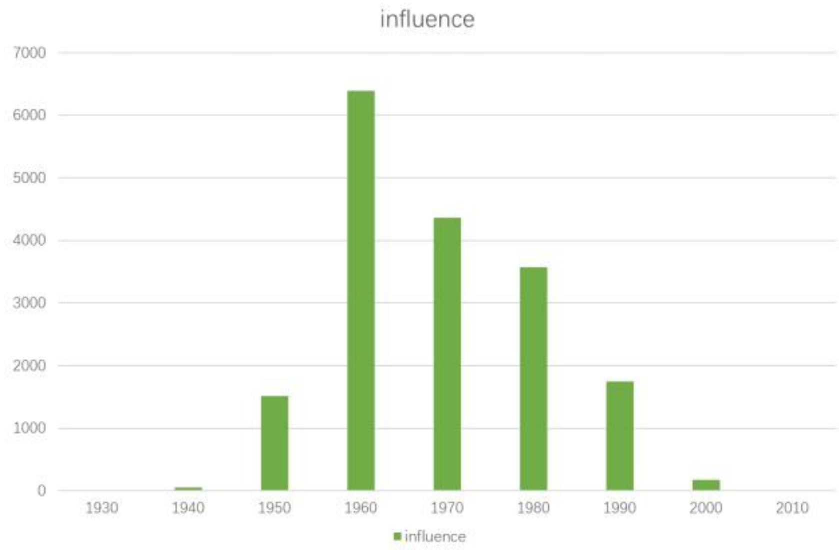

Figure 15: Pop/Rock influence of different eras

It can be clearly seen from the table that the Pop/rock music historically started from 1950s, and the 1960s to the 1990s is the most popular period of pop/Rock music. The out degree of each note represent he number of followers influenced by an artist which represents the indicator that reveal the influence of different artists. 9 of the top 10 most influential artists we found from the our network belong to Pop/Rock, those are The Beatles, Bob Dylan, The Rolling Stones, David Bowie, Led Zeppelin, Jimi Hendrix, The Kinks, The Beach Boys and Black Sabbath, who played an important role in the subsequent development of the Pop/Rock genre.

7.2.2 Influence from other Genres

Find all the nodes in the directed network that have an influence on the development of pop music, and then create a sub-network as shown in the figure below:

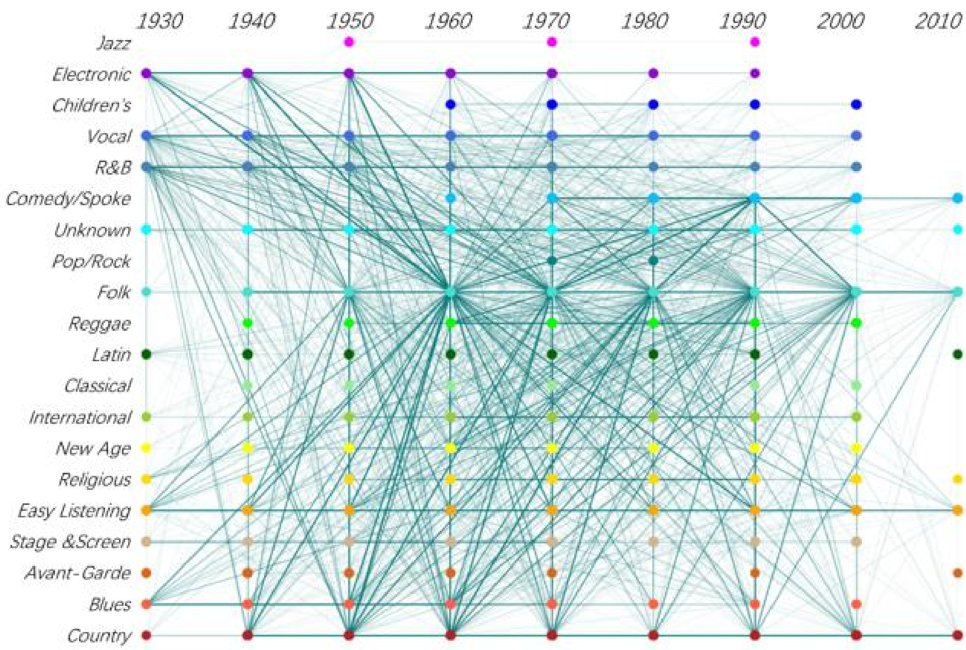

Figure 16: Genres influence Pop/Rock

Analyze the figure, we can find that it was not influenced by many other genres before 1950 year. However, after the development of the 1950s and 1960s, pop rock absorbed numerous music elements from other genres and continued to innovate and develop. Among them, Jazz, Vocal, Blues and country genres have the most significant influence on Pop/Rock.

7.2.3 Characteristic and Popularity

We grouped the data of the three factors obtained in the factor analysis according to music genre and year, averaged each group of data in Pop/Rock, and obtained the following which can explain the trend of Pop/Rock over time.

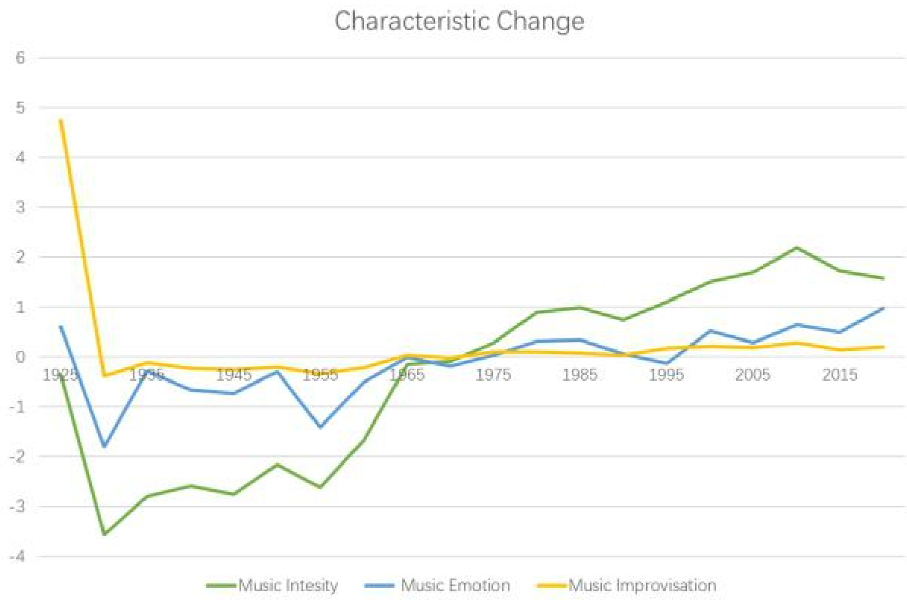

Figure 17: Change of Music Factor

We noticed that both music intensity and music emotion increase over time. The increase of one factor is very obvious, which shows that pop music has more power and rhythm, and it is also famous in the music field for their loud drums, catchy bass lines and powerful vocal hooks.

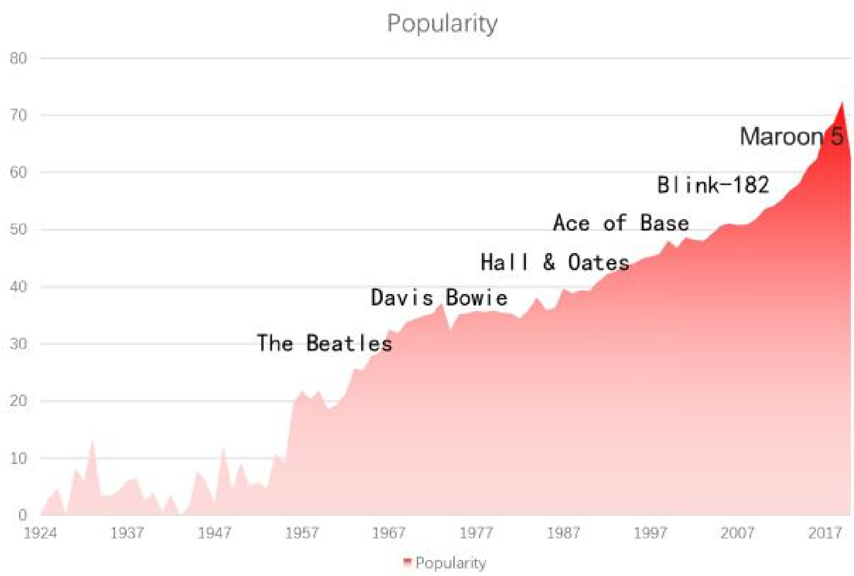

Figure 18: Popularity trend

Through the statistical analysis of popularity data, we can see that Pop/Rock music is constantly prospering, developing and growing. Nowadays, Pop/Rock has become one of the largest music genres.

8 The Influence of Other Factors on Music

- **Politics and society:** At the turn of the century, the two world wars, economic crises and labor-management contradictions have all cast a huge shadow on the psychology of modern people. [10] Reflects the inner feelings of suffering human beings, which is very common in the music works of the early twentieth century. After the end of World War II, cultural society entered the "postmodernist period". The third world countries stepped onto the political arena. The exchanges between civilizations of different nations have become more frequent, and many fusion music works have been born.The society in the second half of the twentieth century was still turbulent. Although the degree of material civilization was high, people did not generally feel happy and comfortable. Rock music, a music genre that expresses dissatisfaction with reality, was born.
- **Technology:** Industrialization reached its peak in the decade after World War II, followed by the arrival of the so-called third wave. Information revolutions, material revolutions, technological revolutions, and biological revolutions all affect all aspects of life. This rapid technological change will also enable artists to continue to innovate in their music works. For example, recording equipment and electronic synthesizers have opened up unlimited space for electronic music.

9 Model Evaluation

9.1 Strengths

- The selection of the network parameters of the passing network is scientific and reasonable considering both time and genre influence.
- The network model is highly applicable convenient which can easily find the relevant information of any node.
- Factor analysis reduces the dimensionality of the data and reduces the computational cost.
- EM Algorithm, compared with K-means, is more general and can form clusters of various sizes and shapes.

9.2 Weaknesses

- The extracted information is not comprehensive. For visualization reasons, only three main factors are extracted in factor analysis, and the original data information cannot completely replace the original data information.
- The clustering results are quite different. Random initialization of the clustering model can sometimes cause large differences in results.

10 A Document

REPORT FOR ICM Society association
REPORTER:ICM TEAM
We have analyzed the music categories, music features and music evolution of different artists, establishing three models, namely graph theory model to analyze the influence of music, multiple factor analysis model including three main factors to judge the classification of music features, and cluster analysis model to analyze the similarity between music. First of all, we build a graph theory model to build a music network to reflect the influence of music. According to the data set, we analyzed the number of influence of each artist, established a map of mutual influence of artists, and obtained the influence of music types. After combining the music publishing time, we get the music influence value and visualize it. Then we summarize the three characteristics of music similarity through principal factor analysis, and cluster all music into five categories to observe the similarity between types. At the same time, we also explore the clustering results of the same type and get the similarity between the types. Through our analysis process, we build the influence model, feature factor model and clustering model, and visualize them. After further analysis, we give relevant examples for more detailed analysis. Finally, we analyze the change of music characteristics, and study the change and development of pop music. We choose one of the types of music, combined with social, political, and technological influence to draw the development conclusion. Our conclusions are as follows: 1. The influence of music can be measured by weight. Different types of music are interrelated 2. The similarity of the same kind of music is stronger than that of different music 3. Each type of music has its own unique determinants, some of which are also contagious 4. The characteristics of music change with time and are influenced by social politics and technology Considering that the data may become more complex in the future, how can our model be improved over time? When we analyze more comprehensive data, our model still works, and we may need to do more detailed analysis. In short, our model has strong practical ability and promotion ability.

In fact, the influence of music and culture is bidirectional. Culture influences the formation of music, and music becomes a part of culture. The influence of music on culture is reflected in the breadth of culture. Music improves people's aesthetic ability, enhances cultural creativity and improves the degree of cultural integration. In order to further study the influence of music on culture, we may need to decide the cultural indicators. However, culture is pluralistic, and the index determines the practical significance of in-depth research.

References

[1]Polykovskiy, D. and Novikov, A., Bayesian Methods for Machine Learning．Coursera and National Research University Higher School of Economics

[2] Rikuto Onishi， Taisuke Otsu， Sample sensitivity for two-step and continuous updating GMM Estimators， Economics Letters 198 (2021) 109685

[3]Jaroslava Hlouskova，Leopold Sögner，GMM estimation of affine term structure models，Econometrics and Statistics 13 (2020) 2 – 15

[4]QinWei，Lin Qingna，Construction of cultural industry development factor model based on factor analysis, artificial intelligence and big data， Microprocessors and Microsystems Volume 82, April 2021, 103880

[5] Jessica Fischer， JiaHea， EckhardKlieme， The structure of teaching practices across countries: A combination of factor analysis and network analysis， Studies in Educational Evaluation Volume 65, June 2020, 100861

[6]Wentao Xiang， Ahmad Karfoul， Chunfeng Yang， Huazhong Shu， RégineLe Bouquin Jeannès， An exact line search scheme to accelerate the EM algorithm: Application to Gaussian mixture models identification，Journal of Computational Science Volume 41, March 2020, 101073

[7]Sergio Medina-González， IoannisGkioulekas， Vivek Dua， Lazaros G.Papageorgiou， A graph theory approach for scenario aggregation for stochastic optimisation， Computers & Chemical Engineering Volume 137, 9 June 2020, 106810

[8]Henrik Ebel， Peter Eberhard， A comparative look at two formation control approaches based on optimization and algebraic graph theory， Robotics and Autonomous Systems Volume 136, February 2021, 103686

[9]A.Ruiz-Frau，A.Ospina-Alvarez，S.Villasante，P.Pita，I.Maya-Jariego，S.de Juan，Using graph theory and social media data to assess cultural ecosystem services in coastal areas: Method development and application，Ecosystem Services Volume 45, October 2020, 101176

[10] Wang Wenxia. The inevitability and reason analysis of the transition from western romantic music to modern music[D]. Hunan Normal University, 2010.

Appendices

Appendix 1
Introduce: Code for the first model, using python

```
####################Graph theory model: directed network###############
import pandas as pd
import numpy as np
import networkx as nx
import matplotlib.pyplot as plt

# create network
network = nx.MultiDiGraph()
for node in nodes:
    network.add_node(node[1], id=node[0], genre=node[2], time=node[3],
                     genre_label=node[4], time_label=node[5])
network.add_weighted_edges_from(weighted_edges, weight='weight')
network['Ben Webster']['Sonny Rollins'][0]['weight']

# calculater out degree denoting the influence of one influencer
out_degree = network.out_degree()
list = []
for i in network.nodes():list.append(out_degree[i])
max_out_degree = max(list)

for node in network.nodes():
network.nodes[node]['coordinates']=(network.nodes[node]['time_label'],network.nodes[node]
['genre_label'])

# set colors for artists from different genre
color_list = ['firebrick', 'tomato', 'chocolate', 'tan', 'orange', 'gold', 'yellow', 'yellowgreen',
'lightgreen','darkgreen','lime', 'turquoise', 'teal', 'cyan', 'deepskyblue', 'steelblue', 'royalblue',
'blue', 'darkviolet','magenta']
for i in network.nodes():
    for j in range(len(color_list)):
        if network.nodes[i]['genre'].find(genre_list[j]) == 0:
            network.nodes[i]['color'] = color_list[j]

    if out_degree[i] >= max_out_degree * .9:
        network.nodes[i]['importance'] = 1500
    elif out_degree[i] >= max_out_degree * .5:
        network.nodes[i]['importance'] = 1000
    else:
```

```
            network.nodes[i]['importance'] = 500
plt.figure(figsize=(20, 15))
nx.draw_networkx_nodes(network,nx.get_node_attributes(network,'coordinates'),node_shape
='.',node_size=[importance for importance in nx.get_node_attributes(network,
'importance').values()],node_color=[color for color in nx.get_node_attributes(network,
'color').values()])
nx.draw_networkx_edges(network,nx.get_node_attributes(network,'coordinates'),edge_color
='gray', alpha=0.1)
plt.show
```

Appendix 2

Introduce: Code for the second modal, using python

```
########################Factor analysis######################
def fa(mydata):
    R = mydata.corr()
    eig_value, eigvector = nlg.eig(R)
    #print(eig_value, eigvector)
    eig = pd.DataFrame()
    eig['names'] = mydata.columns
    eig['eig_value'] = eig_value
    eig.sort_values('eig_value', ascending=False, inplace=True)

    for m in range(1, 10):
        if eig['eig_value'][:m].sum() / eig['eig_value'].sum() >= 0.55:
            s = eig['eig_value'][:m].sum()/ eig['eig_value'].sum()
    A = np.mat(np.zeros((10, 3)))
    for i in range(3):
        A[:, i] = (math.sqrt(eig_value[i]) * eigvector[:, i]).reshape(10, 1)

    a = pd.DataFrame(A)
    a.columns = ['factor1', 'factor2', 'factor3']
    fa = FactorAnalyzer(3,rotation='varimax')
    fa.fit(mydata)
    var = fa.get_factor_variance()
    for i in range(0,3):
        print(var[i])
    df = fa.loadings_
    df = pd.DataFrame(df)
    df.to_csv('factor loadings.csv')

    fa.get_communalities()
```

```
    fa_t_score = np.dot(np.mat(mydata), np.mat(fa.loadings_))
    B = pd.DataFrame(fa_t_score)
    B.to_csv('factor_score.csv')

return B
#########################Clustring#########################
from sklearn.mixture import GaussianMixture
gmm = GaussianMixture(n_components=5,covariance_type ='diag')
gmm.fit(new_data)
pred = list(gmm.predict(new_data))
```